Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Estimation of Breast Cancer with a Combined Feature Selection Algorithm

K. Geetha

Dean, Academics and Research, JCT College of Engineering and Technology, Coimbatore, India

E-mail: geetha.arulmani@gmail.com

Abstract

Image features are considered as a parametric factor that contains some of the specific information about the given image. In simple terms, a feature can be either a size or resolution or color information of an image. From the observed feature, a computer system can predict the nature of the image same as that of a human's perception. In the beginning, the image processing algorithms utilized the features of the image only for the preprocessing and segmentation kinds of applications. An information regarding the noise ratio is considered for the preprocessing work to estimate the amount of smoothness needed to be given to the image. Similarly, the contrast difference or color difference features are widely employed by the segmentation algorithms. The proposed work aims to prove the efficacy of features on breast cancer image classification process using a multilayer perceptron algorithm. An experimental study is conducted on CBIS-DDSM dataset to estimate the importance of local and global features on breast cancer images.

Keywords: Multilayer perceptron, combined features, cancer estimation, region of interest, breast cancer prediction

1. Introduction

Breast cancer is a kind of cancer that generates an abnormal growth on the cells under the skin. A cancer is estimated by identifying the location of the cell that is affected with abnormal growth. Lobules, ducts, and connective tissue are the three main parts of the breast that are affected directly with cancer cell. The lobules gland is a tissue which is helpful for producing milk and the ducts tissue carries the milk towards the nipple. A connecting tissue is the combined formation of fatty and fibrous tissue that covers the ducts and lobules. Among the three main tissues the lobules and ducts are the most common tissues that are affected

DOI: https://doi.org/10.36548/jiip.2022.3.008
Received: 25.08.2022, received in revised form: 21.09.2022, accepted: 05.10.2022, published: 15.10.2022

with cancer cells [1, 2]. In general, the breast cancer can be commonly categorized into two types as invasive ductal carcinoma and invasive lobular carcinoma.

Invasive ductal carcinoma: The cancer cell that is originated in the ducts and spreads over the outer region of the ducts to reach other parts of the body is termed as invasive ductal carcinoma. In some cases these cancer cell can grow on the same place develop a metastasize cancer [3].

Invasive lobular carcinoma: The cancer cell that is started from the lobules are represented as invasive lobular carcinoma. Same as like of the invasive ductal carcinoma the lobular carcinoma also has the ability to spread over the body and other areas of the breast [4]. Apart from these two breast cancers, the Paget's disease, triple negative breast cancer and inflammatory breast cancer are some of the aggressive cancer models that are less common and rare. A breast cancer can be estimated or predicted by the following diagnosis methods [5].

- Complete blood count analysis
- MRI (Magnetic Resonance Image)
- Mammogram
- CT (Computerized Tomography)
- PET (Positron Emission Tomography)
- Bone scan

The above-mentioned diagnosis methods use different kinds of automation tools for estimating the cancer cells through image or data analytic algorithms. The data analytic algorithms are designed based on the logical operations and in recent days the neural network based data mining algorithms are widely employed for such analysis. Similarly the computer assisted image processing algorithms are widely used for before a decade for such prediction but that consumes large amount of computational time. In order to reduce such computational time with an improved accuracy, the deep learning algorithms were introduced in the recent years [6, 7]. The following section explores the attainments and limitations of the deep learning algorithms on estimating the breast cancers.

2. Literature Survey

An efficient Adaboost algorithm was incorporated with CNN to predict the breast cancer in an imbalance dataset. The experimental work indicated a dice co-efficient ratio of

96.4% on 5000 images. Similarly the work gives a Matthews correlation coefficient of 98.5% [1]. A breast cancer detection technique was proposed to estimate the breast cancers on histology images. The work utilized the CNN algorithm on bioimaging2015 dataset with a clustering model. The experimental outcome indicated a prediction accuracy of 88.89% and that is far better than the previous Araujo model verified with the same dataset on 78% accuracy [2]. A hybrid deep neural network approach was structured to classify the cancer cells from a histopathological image consists of 3771 images. The experimental analysis produced an outcome of 91.3% on 4 classes of images. The work was also compared with the VGG16 and ResNet-50 algorithm that attained the accuracy rate of 79.2 and 81.6 percentages respectively [3].

A weight transfer algorithm was proposed to change pre-trained weights of DenseNet121 and ResNet50 in the imagenet dataset. The work reported an accuracy of 100% on binary classification and 98% on multi-classification [4]. A deep learning based optimization method was designed to generate an automated cancer detection model. The work was organized to predict the 20 stages of breast cancer growth in mammographic images. The experiment report indicated a sensitivity of 96% and specificity of 93% [5].

A deep neural network based multi-classification approach was designed for predicting breast cancers in BreakHis dataset. The work designed a biopsy microscopic image cancer network that has the ability to estimate 8 classes of breast cancers with 95.48% of accuracy [6]. A patch based deep learning network was structured to classify breast cancers from histopathological Images. The algorithm utilized the deep belief network for an analysis in DRYAD dataset and that gives an outcome of 86% accuracy [7].

A multi-class SVM approach was framed to classify the breast cancers from mammogram images. An experiment was performed in the work at Mini-MIAS dataset and that produces an accuracy of 96.9% where as in fine KNN the accuracy is 94.8% [8]. An incremental boosting algorithm was designed to estimate the breast cancers from histology images. The experimental work was performed in BreakHis dataset and Bioimaging 2015 dataset, and that produced accuracies of 96.3% and 98.9% respectively [9]. A patch correction approach was enforced to the CNN algorithm for improving its classification accuracy on breast cancer detection application. The performance of the developed algorithm was also compared with the SVM algorithm contains the same set of features. The work was performed in the BreakHis dataset and attains an accuracy difference of 1% between the verified algorithms [10].

3. Proposed Work

The literature section indicates the outcome of the previous algorithms on different datasets. It has been found that the existing algorithms were solely trust upon the classification algorithms that are used for the analysis. The proposed work tries to explore the importance of the feature selection process extracted from the dataset images on classification process. To do that the dataset images were extracted with the information on local and global features. The combined feature algorithm represents the features that are taken from both local and global feature process. The workflow of the proposed model is represented in figure 1.

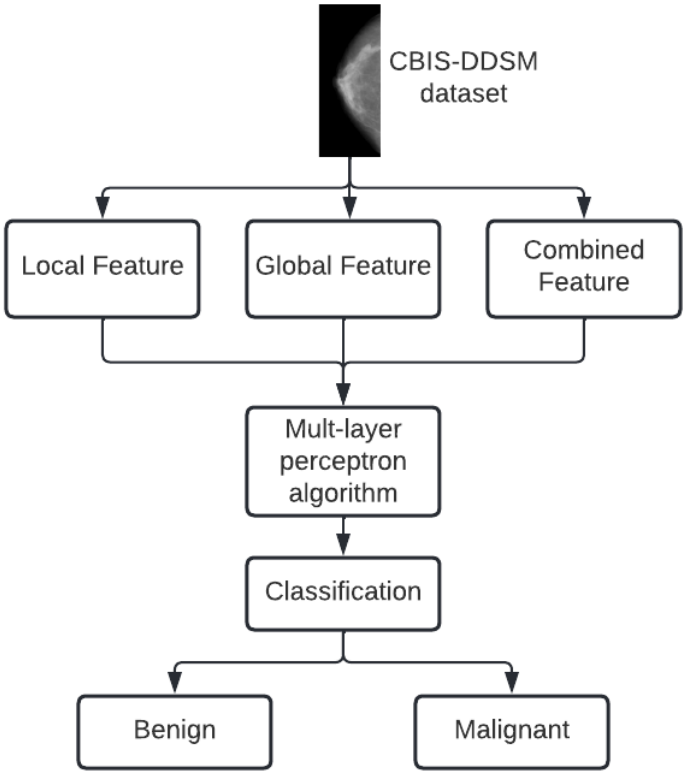

Figure 1. Workflow of the proposed approach

3.1 Local Feature

The local features are the collection of data from the given image on some specifies area or region in the image. The local features are widely applied to the medical images that are applied with a segmentation algorithm. The efficacy of local features can clearly estimate the data information that is available in such local segmented region. The performance of the local feature is generally good when the local portion is comparatively differentiated from its surroundings.

3.2 Global Feature

The global features are the collection of information on the entire category. However, this improves the computational complexity due to the addition of large feature information. The global features are widely included in the multi-class classification models as the features are different on each class of images. Figure 2 represent an overview of the different types of feature extraction models.

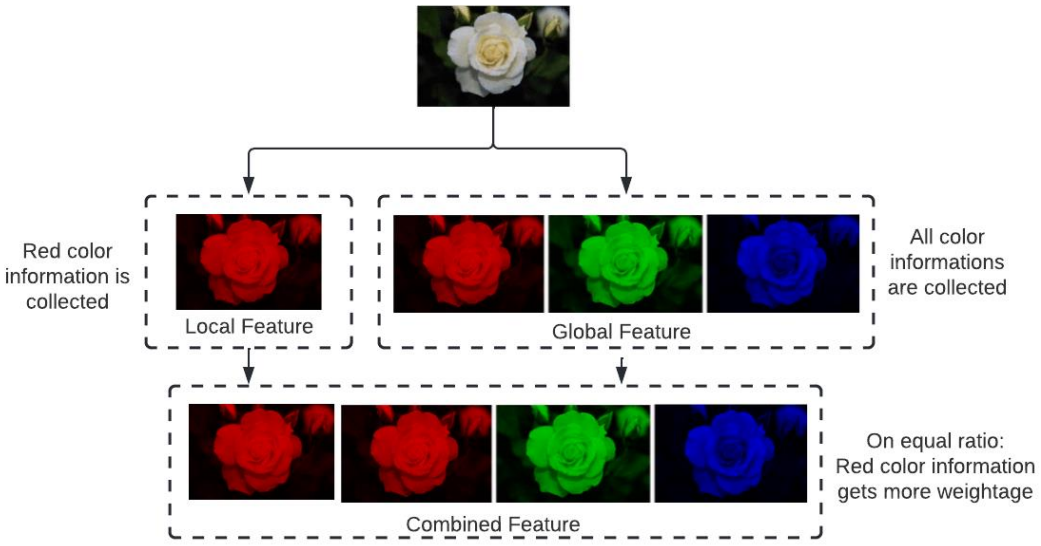

Figure 2. General view of different feature extraction models

3.3 Combined Feature

The combined feature is the process of sharing both local and global feature information to the neural network algorithm. The combined feature extraction technique is included in the applications that require more attention towards some specific tasks. In general the combined features are not high in computation as it has only few more features in its consideration than the global features.

3.4 Multilayer Perceptron (MLP)

The architectural view of an MLP algorithm looks almost similar as of a feed forward neural network algorithm that contains multiple layers of perceptron with an activation function. The counts of input and output layers are same in MLP network but it contains more than one hidden layer in its workflow. The architectural overview of an MLP is shown in figure 3. The input information that is extracted from the given image is taken into consideration by the MLP algorithm through its input layer. The input layer neurons are

connected towards the hidden layer in a single direction. However, the operational weightage between the input and hidden layer can be operated with a tuning parameter for improving its betterment. The MLP also includes certain activation functions including ReLUs, tanh and sigmoid for its analysis. The correlation difference observed between the two classes of images in the activation function explores the accuracy outcome of the MLP.

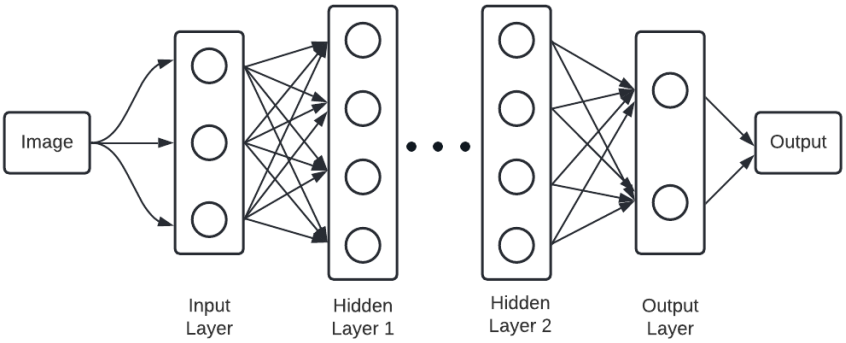

Figure 3. Architectural view of MLP model

4. Experimental Analysis

The performance of the proposed work is verified with CBIS-DDSM [11] dataset that contains the images of normal, benign and malignant category of breast cancer. The proposed work considers only the benign and malignant category for its operation. In total 912 images are available for benign category and 784 images are available for the malignant category. The dataset images are segregated with 60:40 ratio for the training and testing analysis. The experimental work was performed in the TensorFlow platform on Windows 10 operating system contains 64GB RAM for the analysis.

Table 1. Experimental outcome of the proposed approach

Model	Accuracy	F1score	AUC
MLP + Local Features	0.912	0.901	0.914
MLP + Global Features	0.863	0.85	0.87
MLP + Combined Features	0.89	0.867	0.889

Table 1 explores the outcome of the proposed work on its 3 different kinds of feature selection approaches. The experimental work finds the outcome of the proposed work in terms of accuracy, F1score and Area Under the Curve (AUC).

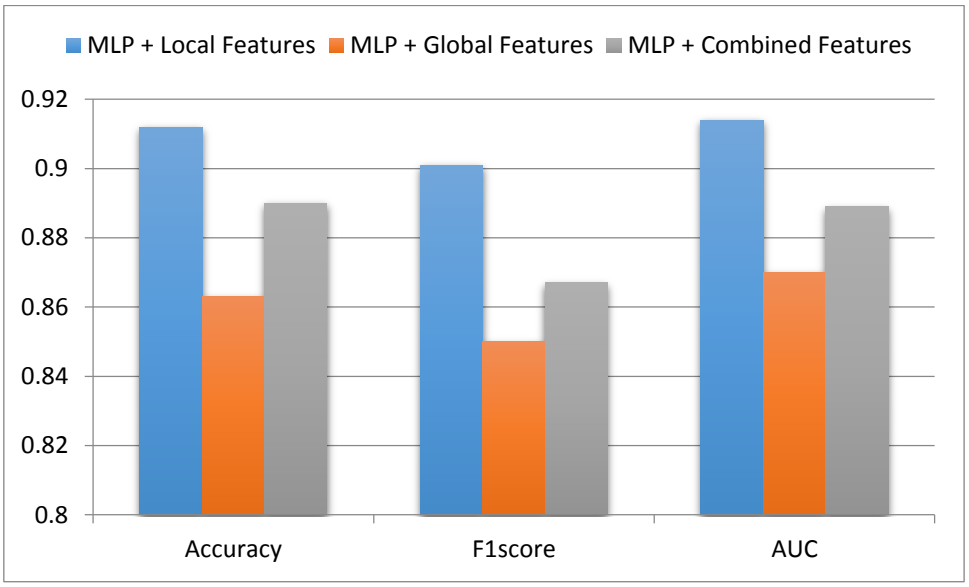

Figure 4. Comparative analysis of the proposed work

Figure 4 indicates that the MLP algorithm combined with local feature gives a better outcome on all the parameters in the analysis. Similarly, the combined feature category comes next to the local feature as it carries more ratio of information on the local features. The outcome of the global feature model indicates a poor outcome since there is no necessity for considering all the feature parameters.

5. Conclusion

A breast cancer detection approach is developed in the work to estimate the images' category i.e., benign or malignant. The work equips a Multilayer Perceptron algorithm for the classification process and its efficacy is verified by giving different types of feature selection approaches. The dataset images were not considered in the work for the preprocessing step, and so the information is directly extracted from the raw image. The experimental analysis indicates that the proposed MLP algorithm performs better in terms of utilizing the local feature for its operation. In general, the medical images perform better with local features and in rare cases, the performances are found better with global and combined features. The work analysis determines that the CBIS-DDSM dataset performs better, attaining 91.2% accuracy with the local features.

References

[1] Fahad Ullah, Mohammad. "Breast cancer: current perspectives on the disease status." Breast Cancer Metastasis and Drug Resistance (2019): 51-64.

[2] Yedjou, Clement G., Jennifer N. Sims, Lucio Miele, Felicite Noubissi, Leroy Lowe, Duber D. Fonseca, Richard A. Alo, Marinelle Payton, and Paul B. Tchounwou. "Health and racial disparity in breast cancer." Breast cancer metastasis and drug resistance (2019): 31-49.

[3] Celik, Yusuf, Muhammed Talo, Ozal Yildirim, Murat Karabatak, and U. Rajendra Acharya. "Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images." Pattern Recognition Letters 133 (2020): 232-239.

[4] Mouabbi, Jason A., Amy Hassan, Bora Lim, Gabriel N. Hortobagyi, Debasish Tripathy, and Rachel M. Layman. "Invasive lobular carcinoma: an understudied emergent subtype of breast cancer." Breast Cancer Research and Treatment (2022): 1-12.

[5] Jiménez-Gaona, Yuliana, María José Rodríguez-Álvarez, and Vasudevan Lakshminarayanan. "Deep-learning-based computer-aided systems for breast cancer imaging: a critical review." Applied Sciences 10, no. 22 (2020): 8298.

[6] Han, Zhongyi, Benzheng Wei, Yuanjie Zheng, Yilong Yin, Kejian Li, and Shuo Li. "Breast cancer multi-classification from histopathological images with structured deep learning model." Scientific reports 7, no. 1 (2017): 1-10.

[7] Wei, Benzheng, Zhongyi Han, Xueying He, and Yilong Yin. "Deep learning model based breast cancer histopathological image classification." In 2017 IEEE 2nd international conference on cloud computing and big data analysis (ICCCBDA), pp. 348-353. IEEE, 2017.

[8] Zheng, Jing, Denan Lin, Zhongjun Gao, Shuang Wang, Mingjie He, and Jipeng Fan. "Deep learning assisted efficient AdaBoost algorithm for breast cancer detection and early diagnosis." IEEE Access 8 (2020): 96946-96954.

[9] Li, Yuqian, Junmin Wu, and Qisong Wu. "Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning." IEEE Access 7 (2019): 21400-21408.

[10] Yan, Rui, Fei Ren, Zihao Wang, Lihua Wang, Tong Zhang, Yudong Liu, Xiaosong Rao, Chunhou Zheng, and Fa Zhang. "Breast cancer histopathological image classification using a hybrid deep neural network." Methods 173 (2020): 52-60.

[11] Yari, Yasin, Thuy V. Nguyen, and Hieu T. Nguyen. "Deep learning applied for histological diagnosis of breast cancer." IEEE Access 8 (2020): 162432-162448.

[12] Sha, Zijun, Lin Hu, and Babak Daneshvar Rouyendegh. "Deep learning and optimization algorithms for automatic breast cancer detection." International Journal of Imaging Systems and Technology 30, no. 2 (2020): 495-506.

[13] Murtaza, Ghulam, Liyana Shuib, Ghulam Mujtaba, and Ghulam Raza. "Breast cancer multi-classification through deep neural network and hierarchical classification approach." Multimedia Tools and Applications 79, no. 21 (2020): 15481-15511.

[14] Hirra, Irum, Mubashir Ahmad, Ayaz Hussain, M. Usman Ashraf, Iftikhar Ahmed Saeed, Syed Furqan Qadri, Ahmed M. Alghamdi, and Ahmed S. Alfakeeh. "Breast cancer classification from histopathological images using patch-based deep learning modeling." IEEE Access 9 (2021): 24273-24287.

[15] Kaur, Prabhpreet, Gurvinder Singh, and Parminder Kaur. "Intellectual detection and validation of automated mammogram breast cancer images by multi-class SVM using deep learning classification." Informatics in Medicine Unlocked 16 (2019): 100151.

[16] Vo, Duc My, Ngoc-Quang Nguyen, and Sang-Woong Lee. "Classification of breast cancer histology images using incremental boosting convolution networks." Information Sciences 482 (2019): 123-138.

[17] Ahmad, Nouman, Sohail Asghar, and Saira Andleeb Gillani. "Transfer learning-assisted multi-resolution breast cancer histopathological images classification." The Visual Computer 38, no. 8 (2022): 2751-2770.

[18] Lee, Rebecca Sawyer, Francisco Gimenez, Assaf Hoogi, Kanae Kawai Miyake, Mia Gorovoy, and Daniel L. Rubin. "A curated mammography data set for use in computer-aided detection and diagnosis research." Scientific data 4, no. 1 (2017): 1-9.

Author's biography

K. Geetha is presently working as a Dean-Academics and Research at JCT College of Engineering and Technology, Coimbatore, India. Her area of research includes image processing, computer vision, artificial intelligence, optimization techniques, power systems, process control and automations.